Pier Francesco Perri (University of Calabria, Italy)

Eliciting Sensitive Data by Indirect Questioning Techniques: Some Recent Applications and Methodological Advances

Topic 2 – Learning more from what we already know

Keywords: sensitive questions, privacy protection, randomized response theory, item sum technique

Introduction

In many fields of the applied research, mostly in sociological, economics, demographic, ecological and medical studies, the investigator very often has to gather information concerning highly personal, sensitive, stigmatizing, and perhaps incriminating issues such as drug addiction, domestic violence, racial prejudice, illegal income, noncompliance with laws and regulations.

Doing research on delicate themes - especially by Direct Questioning (DQ) modes - is not an easy matter since it is likely to meet with three source of errors: (1) refusal to cooperate (unit-non-response); (2) refusal to answer specific questions (item-non-response); (3) untruthful answers (measurement error). These errors can seriously flaw the quality of the data and, thus, jeopardize the usefulness of the collected information for subsequent analyses including inference on unknown characteristics of the population under study.

Although these errors cannot be totally avoided, they may be mitigated by increasing respondent cooperation, carefully considering some key points such as the modes the survey is administrated, the format of the questionnaire, the wording and the placing of the sensitive items in the questionnaire.

Alternatively, Indirect Questioning Techniques (IQTs) could be used (Chaudhuri 2011; Chaudhuri and Christofides, 2013). Here we focus on some applications of the Randomized Response Theory (RRT) and on some methodological advances for the Item Sum Technique.

Methods / Problem statement

Dishonest or misleading answers generate in sensitive surveys the so called “social desirability bias” which refers to the tendency to present oneself on a positive light. Survey participants exhibit this bias when they overreport socially acceptable attitudes or behaviors (e.g., giving alms) and underreport socially disapproved behaviors (e.g, earing illegal income).

This type of bias generally produces an over- or under-estimation of the behavior under study which may lead to inconsistent analyses and wrong conclusions. Since the sixties, a variety of questioning methods - different from the conventional survey modes - have been devised to reduce nonresponse, to ensure respondent anonymity and cutting down evasive answers and underreporting/overreporting of socially undesirable/acceptable acts.

These methods are generally known as IQTs and obey the principle that no direct questions are posed to the survey participants and, thus, there is no need for the respondents to openly reveal if they are actually engaged in sensitive behaviors. In this way, privacy is protected since answers remain confidential to the respondents and, consequently, their true status remains uncertain and undisclosed to the researcher. The IQTs represent a category of strategies for eliciting sensitive information which encompasses various approaches: the RRT holds a prominent position certainly.

Results / Proposed solution

Roughly speaking, the RRT adopts a randomization device (decks of cards, colored numbered balls, dice, coins, etc.) to hide the true answer in the sense that the interviewees reply to one of two or more selected questions depending on the result of the randomizer.

The technique has been widely employed in many researches. For instance, recently Perri et al. (2015) employed the “crossed model” proposed by Lee at al. (2013) in a pilot survey (868 women) conducted in Calabria (Italy) to investigate the prevalence of two sensitive characteristics, say induced abortion among foreign women residing in the region and irregular immigrant status.

Collecting data on these two attributes by means of DQ modes produces underestimates of the diffusion of the phenomena due to the stigmatizing nature of the investigated topics. The effectiveness of the crossed model has ben also investigated in another study (289 individuals) to simultaneously estimate the prevalence of individuals who smoke cannabis and support cannabis legalization. The crossed model requires two decks of cards to drive the answers.

In general, using physical devices limits the application of the RRT exclusively to face-to-face personal interviews and may also be a little time consuming and costly than DQ modes. To overcome these drawbacks, in a self-administered questionnaire survey (680 workers) concerning the estimation of the prevalence of workplace mobbing, the triangular model (Tian and Tang 2014) has bee

Conclusions

During the conference, the survey plan and the results of the studies previously mentioned will be discussed with a focus on the “more-is-better” principle according to higher RRT prevalence estimates than DQ estimates are observed.

Comments on the feasibility and the practical implementation of the randomization methods will also be provided.

Finally, some advances on the Item Sum Technique will be mentioned in order to motivate and stimulate the use of the technique in real analyses. In particular, by means of a simulation study, it will be shown how to improve the estimates of a sensitive mean when auxiliary variables on the population under study are available and used at the design and estimation stages.